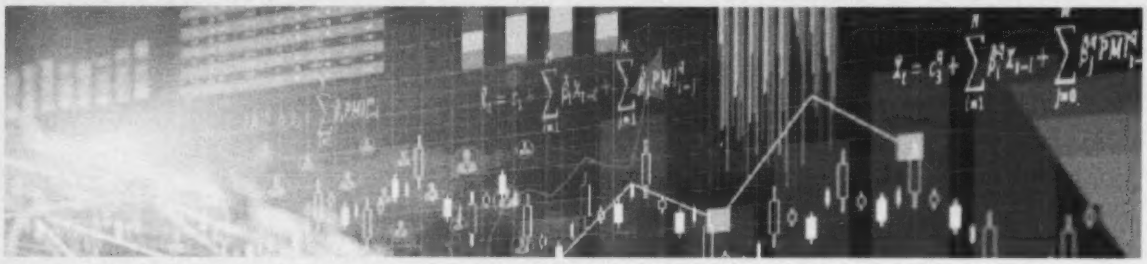
BANK OF CANADA
BANQUE DU CANADA

Staff Working Paper/Document de travail du personnel 2016-18

A General Approach to Recovering Market Expectations from Futures Prices with an Application to Crude Oil



by Christiane Baumeister and Lutz Kilian

Bank of Canada staff working papers provide a forum for staff to publish work-in-progress research independently from the Bank's Governing Council. This research may support or challenge prevailing policy orthodoxy. Therefore, the views expressed in this paper are solely those of the authors and may differ from official Bank of Canada views. No responsibility for them should be attributed to the Bank.

www.bank-banque-canada.ca

Bank of Canada Staff Working Paper 2016-18

April 2016

A General Approach to Recovering Market Expectations from Futures Prices with an Application to Crude Oil

by

Christiane Baumeister[1] and Lutz Kilian[2]

[1]Department of Economics
University of Notre Dame
CEPR
cjsbaumeister@gmail.com

[2]Department of Economics
University of Michigan
CEPR
Corresponding author: lkilian@umich.edu

ISSN 1701-9397

© 2016 Bank of Canada

Acknowledgements

We thank Cynthia Wu for providing access to the code used in Hamilton and Wu (2014). We have benefited from helpful discussions with Reinhard Ellwanger, Refet Gürkaynak, James Hamilton, Matthew Kahn, Ryan Kellogg, Stefan Nagel, Michel Robe, Allan Timmermann and Cynthia Wu. We thank David Finer for excellent research assistance. Lutz Kilian acknowledges financial support from the Bank of Canada.

Abstract

Futures markets are a potentially valuable source of information about price expectations. Exploiting this information has proved difficult in practice, because time-varying risk premia often render the futures price a poor measure of the market expectation of the price of the underlying asset. Although this expectation in principle may be recovered by adjusting the futures price by the estimated risk premium, a common problem is that there are as many measures of the market expectation as there are estimates of the risk premium. We propose a general solution to this problem that allows us to select the most accurate estimate of the expectation for any set of risk premium estimates. We illustrate this approach by solving the long-standing problem of how to estimate the market expectation of the price of crude oil. We provide a new measure of oil price expectations that is substantially more accurate than the alternatives and more economically plausible. Our analysis has implications for the estimation of economic models of energy-intensive durables, for oil price forecasting and for the measurement of oil price shocks.

JEL classification: C53, D84, G14, Q43
Bank classification: Econometric and statistical methods; International topics

Résumé

Les marchés des contrats à terme peuvent constituer une source d'information précieuse au sujet des anticipations de prix. Toutefois, dans la pratique, il s'avère difficile d'exploiter cette information, les prix des contrats à terme offrant souvent une mesure médiocre des attentes du marché à l'égard des prix des actifs sous-jacents en raison de la présence de primes de risque variables dans le temps. Bien qu'il soit possible en principe de calculer ces attentes en ajustant les prix des contrats à terme en fonction de la prime de risque estimée, le fait qu'il existe autant de mesures des anticipations du marché que d'estimations de la prime de risque représente un problème courant. À ce problème, les auteurs proposent une solution générale qui leur permet de sélectionner, pour tout ensemble d'estimations de la prime de risque, l'estimation la plus exacte des anticipations. Pour illustrer cette méthode, ils résolvent le problème persistant de l'évaluation des attentes du marché à l'égard des prix du pétrole brut. La nouvelle mesure des anticipations concernant les prix du pétrole présentée par les auteurs est beaucoup plus précise que les autres mesures et plus plausible du point de vue économique. Leur analyse a des incidences sur l'estimation des modèles économiques relatifs aux biens durables à forte intensité énergétique, sur les prévisions des prix du pétrole et sur l'évaluation des chocs pétroliers.

Classification JEL : C53, D84, G14, Q43
Classification de la Banque : Méthodes économétriques et statistiques; Questions internationales

Non-Technical Summary

Many researchers have treated the futures price as a measure of the market expectation of the spot price. To the extent that there is a time-varying risk premium in futures markets, however, the futures price will deviate from the market expectation of the price of the underlying asset. This risk premium in principle can be estimated, but, depending on the model specification, the risk premium estimates may differ substantially. Because in practice there are as many measures of the market expectation as there are estimates of the risk premium, the problem of how to recover the market expectation has remained unresolved so far. We propose a general solution to this problem that allows us to select the best possible estimate of the market expectation for any set of risk premium estimates. We illustrate the usefulness of this approach by solving the long-standing problem of how to estimate the market expectation of the price of crude oil.

We start by examining a wide range of methods of estimating the time-varying risk premium in oil futures markets and by demonstrating that alternative estimates of the risk premium for the same month may differ by as much as $61. This model uncertainty makes it impossible to reach a consensus on the expectation of the price of crude oil without further analysis. We show that the most accurate estimate of the expectation of the price of oil is obtained by adjusting the oil futures price by the risk premium inferred from the term-structure model of Hamilton and Wu (2014). This expectations measure not only is statistically more accurate than any other estimate of the oil price expectation that we considered (including the unadjusted oil futures price), but is economically more plausible. Accounting for the risk premium also helps resolve the puzzle that oil futures markets seemingly failed to predict the sustained surge in the price of oil between 2003 and mid-2008. Our analysis has important implications for a wide range of macroeconomic and microeconomic models.

1. Introduction

Price expectations play a key role in a wide range of forward-looking economic models. One source of information about price expectations is futures markets. Many researchers have treated the futures price as a measure of the market expectation of the spot price. To the extent that there is a time-varying risk premium in futures markets, however, the futures price will deviate from the market expectation of the price of the underlying asset. This risk premium in principle can be estimated, but, depending on the model specification, the risk premium estimates may differ substantially. Because in practice there are as many measures of the market expectation as there are estimates of the risk premium, the problem of how to recover the market expectation has remained unresolved so far. We propose a general solution to this problem that allows us to select the best possible estimate of the market expectation for any set of risk premium estimates. The central idea is that – in the presence of a risk premium – the risk-premium adjusted futures price is the conditional expectation of the price and hence minimizes the mean-squared prediction error (MSPE) (e.g., Granger 1969). This theoretical result allows one to rank alternative model specifications based on their MSPE and to identify the most accurate measure of expectations.

We illustrate the usefulness of this approach by solving the long-standing problem of how to estimate the market expectation of the price of crude oil. We start by examining a wide range of methods of estimating the time-varying risk premium in oil futures markets and by demonstrating that alternative estimates of the risk premium for the same month may differ by as much as $61. This model uncertainty makes it impossible to reach a consensus on the expectation of the price of oil without further analysis. We then evaluate the merits of these estimates based on the predictive content of the implied oil price expectations. Based on the

MSPE ranking, the most accurate estimate of the expectation of the price of oil is obtained by adjusting the oil futures price by the risk premium inferred from the term-structure model of Hamilton and Wu (2014). This expectations measure not only is more accurate than any other estimate of the oil price expectation that we considered (including the unadjusted oil futures price), but also is economically more plausible.

Our new measure of oil price expectations consisting of monthly time series of expectations at horizons of up to one year can be used in estimating economic models of automobile purchases, investment decisions under uncertainty, environmental policies and regulatory reforms. It also allows a retrospective analysis of the evolution of the market's oil price expectations since 1992, and it provides an important input for empirical models of learning in financial markets. We find, for example, that the futures market did not anticipate the oil price decline of late 2008, but that market expectations were substantially lower than the 12-month oil futures price in mid-2008. Another interesting finding is that between 2010 and mid-2013 the one-year-ahead market expectation of the price of oil had stabilized near $90, notwithstanding the fluctuations in the oil futures price. We also show that the systematic failure of the oil futures price as a predictor of the spot price between 2003 and mid-2008 can be largely explained by fluctuations in the risk premium.

Oil price expectations also play an important role in the macroeconomic projections generated by central banks and international organizations. It is common for central banks to rely on oil futures prices in forming out-of-sample oil price forecasts. Given the empirical success of the Hamilton-Wu term structure model in estimating historical oil price expectations, a natural question is whether this method could also be used to improve the accuracy of central banks' oil price expectations. We provide some tentative evidence that risk-premium adjusted oil futures

prices obtained from real-time estimates of the Hamilton-Wu model at horizons up to nine months are more accurate out of sample than the unadjusted oil futures price.

The remainder of the paper is organized as follows. Section 2 reviews the emerging consensus on the existence of a time-varying risk premium in oil futures markets and explains why the existing evidence is inconclusive. In section 3 we demonstrate how the time-varying risk premium in oil futures markets may be estimated using a novel regression approach, and how these estimates can be used to measure the market's expectations of the dollar price of crude oil at selected horizons. In section 4 we quantify the historical evolution of alternative risk premium estimates and their dispersion across model specifications. Section 5 introduces our procedure for ranking alternative estimates of the time-varying risk premium, explains its theoretical rationale, and illustrates its implementation. Section 6 highlights what we learn about the market's oil price expectations from the preferred model specification. Section 7 examines the out-of-sample accuracy of risk-premium adjusted oil futures prices. Section 8 summarizes additional results based on quarterly data contained in the appendix. The concluding remarks are in section 9.

2. Risk premia in the oil futures market: What we know and why we care

For many years, the standard practice among policymakers and central bankers, in the business community, in the financial press and in the academic literature, has been to interpret the price of West Texas Intermediate (WTI) crude oil futures as the market expectation of the spot price of WTI crude oil. For example, there is a long tradition of using oil futures prices as proxies for energy price expectations in empirical models of the purchases of energy-intensive durables, in models of the effect of uncertainty on investment decisions, and in models of the impact of regulatory policies such as automotive fuel standards and gasoline taxes (e.g., Busse, Knittel and

Zettelmeyer 2013; Kellogg 2014; Allcott and Wozny 2014).

This practice has been challenged in recent years by a large number of empirical studies documenting the existence of time-varying risk premia in the oil futures market. Examples include De Roon, Nijman, and Veld (2000), Sadorsky (2002), Pagano and Pisani (2009), Acharya, Lochstoer, and Ramadorai (2013), and Etula (2013). These studies provide direct evidence that returns in oil futures markets can be predicted using a range of aggregate and commodity-market specific financial and macroeconomic variables, which implies the presence of a time-varying risk premium.[1] Singleton (2014) summarizes this new consensus by noting that "the evidence for time-varying risk premiums in oil markets … seems compelling". This risk premium drives a wedge between the oil futures price and the expected spot price of crude oil, invalidating the use of oil futures prices as measures of market expectations.

Although the evidence in this literature may seem overwhelming at first sight, closer inspection reveals that it is difficult to draw general conclusions from these empirical studies. First, sometimes in this debate little distinction is made between results for crude oil and for other commodities.[2] A closely related problem is that many studies do not explicitly focus on crude oil, but estimate time-varying risk premia for portfolios of several energy commodities. Such studies are not informative about the question of time-varying risk premia in the crude oil market.[3] Second, studies focusing on crude oil tend to differ greatly in the sample period covered and in the horizon for which the risk premium is computed, making it difficult to compare the

[1] The month-to-month return on futures is typically defined as $\left(F_{t+1}^{h-1} - F_t^h\right)/F_t^h$, where F_t^h denotes the price of a futures contract with a maturity of h months entered into in month t.

[2] For example, Singleton (2014) cites Fama and French (1987) as having provided evidence of a time-varying risk premium in the crude oil market, yet oil was never considered in their paper. At the time Fama and French (1987) was written, the WTI oil futures market was still in its infancy.

[3] A case in point is the study by Hong and Yogo (2012), which provides results for an energy portfolio consisting of heating oil, gasoline, crude oil, natural gas, and propane gas, but no results that speak directly to the question of the risk premium in the crude oil market.

estimates. Third, there is no agreement on the question of which predictors should be used for estimating the risk premium. Whereas one study might favor one set of predictors, the next study may focus on an entirely different set of predictors. This fact suggests that there is scope for improving risk premium estimates by combining predictors.

Our paper provides a systematic investigation of the evidence for a time-varying risk premium in the oil futures market with the objective of quantifying its evolution in dollar terms and of deriving the implied oil price expectations. Because traditional return regressions based on holding-period returns do not yield estimates of the magnitude of the dollar risk premium at the horizon of interest, as required for the construction of measures of the implied oil price expectations, we instead estimate the risk premium as the predictable component of the payoff of an oil futures contract. We evaluate the predictive content of a wide range of predictors, drawing on the existing literature on time-varying risk premia in commodity futures markets. When the original study did not consider crude oil specifically we extend the analysis to the oil futures market. We also consider predictors that heretofore have only been used in modelling returns on energy portfolios, and we generalize existing model specifications by combining many predictors. For each set of predictors we follow as closely as possible the data definitions in the literature, but we extend the sample until June 2014, which in many cases amounts to a substantial increase in the sample size. We present results for prediction horizons of 3, 6, 9, and 12 months, and we evaluate all model specifications based on the same evaluation period to the extent that the data are available.[4]

[4] Restricting the horizon to 12 months allows a systematic evaluation of the accuracy of the predictive models over a long evaluation period. Clearly, for some applications longer maturities are of interest. For example, Kellogg (2014) focuses on the 18-month futures contract. It would be straightforward to extend our analysis to such horizons, but only at the cost of a shorter evaluation period.

3. Estimating the risk premium and the expectation of the price of oil

Oil futures markets facilitate the transfer of non-diversifiable risk to market participants who are willing to bear it. The market price of this risk is known as the risk premium. Let F_t^h denote the current dollar price of an oil futures contract maturing h periods from now, and S_{t+h} the corresponding spot price of oil at the expiry date of the futures contract. The random payoff of a long position is $S_{t+h} - F_t^h$. Standard no arbitrage arguments imply that

$$E_t\left[Q_{t+h}\left(S_{t+h} - F_t^h\right)\right] = 0,$$

where Q_{t+h} denotes a stochastic pricing kernel. Rearranging this expression yields

$$F_t^h = E_t\left[S_{t+h}\right] + \text{cov}\left(S_{t+h}, Q_{t+h}\right) / E_t\left[Q_{t+h}\right]$$

where $\text{cov}\left(S_{t+h}, Q_{t+h}\right) / E_t\left[Q_{t+h}\right]$ is the risk premium. It is readily apparent that $F_t^h = E_t\left[S_{t+h}\right]$ if and only if $\text{cov}\left(S_{t+h}, Q_{t+h}\right) / E_t\left[Q_{t+h}\right] = 0$. In that case, $E_t\left[S_{t+h} - F_t^h\right] = 0$, where $S_{t+h} - F_t^h$ denotes the error from predicting the spot price of oil based on the oil futures price. Thus, evidence of a predictable component in the prediction error such that $E_t\left[S_{t+h} - F_t^h\right] \neq 0$ would be consistent with the presence of a time-varying risk premium.

The prediction error $S_{t+h} - F_t^h$ is not stationary and must be transformed to allow us to estimate its predictable component by regression methods. For this purpose it is useful to normalize the dependent variable without loss of generality as $\left(S_{t+h} - F_t^h\right) / F_t^h$. This variable may be viewed as the payoff on an oil futures contract held to maturity. We use the exact percent deviation rather than its log approximation because the percent deviations can be large. The risk premium then may be estimated from the regression

$$\frac{S_{t+h}-F_t^h}{F_t^h}=\alpha_h+\beta_h x_t+\nu_{t+h}, \tag{1}$$

where x_t denotes the vector of predictors, ν_{t+h} is the mean zero error term, and α_h and β_h are the horizon-specific regression coefficients. The predictors included in x_t are chosen based on the existing literature on the time-varying risk premium in oil futures markets. Solving equation (1) for S_{t+h} yields

$$S_{t+h}=\left(1+\alpha_h+\beta_h x_t+\nu_{t+h}\right)F_t^h.$$

Hence, the market's expectation of the spot price of oil h months from now

$$E_t\left(S_{t+h}\right)=F_t^h\left(1+\alpha_h+\beta_h x_t\right), \tag{2}$$

may be constructed as the risk-premium adjusted oil futures price, where the dollar risk premium is $RP_t^h=F_t^h-E_t(S_{t+h})$. The OLS estimator of regression model (1) generates a consistent estimate of RP_t^h and of the implied oil price expectation. Moreover, using the full sample in estimating this regression under the maintained assumption of stationarity will result in optimal estimates of the risk premium at date t and hence of the oil price expectation prevailing in the market at that point in time.[5]

4. Alternative estimates of the time-varying risk premium

Our analysis focuses on the WTI price of crude oil.[6] We consider three classes of models for estimating the time-varying risk premium. The first class of models follows Fama and French

---

[5] It might seem that an alternative approach would have been to standardize the prediction error as $\left(F_t^h-S_{t+h}\right)/S_{t+h}$. Although the latter transformation also renders the dependent variable stationary, it would not be useful for our purposes, because $E_t\left[\left(F_t^h-S_{t+h}\right)/S_{t+h}\right]=0$ does not imply $E_t\left[F_t^h-S_{t+h}\right]=0$ due to Jensen's inequality. Because $E_t\left[\left(F_t^h-S_{t+h}\right)/S_{t+h}\right]$ will be positive in expectation and time varying, when $E_t\left[F_t^h-S_{t+h}\right]=0$, evidence that this dependent variable is predictable need not be evidence of a time-varying risk premium.

[6] For a related analysis of the Brent price of crude oil see Baumeister and Kilian (2016a, b).

(1987, 1988) in focusing on the predictive power of the basis, defined as $\left(F_t^h - S_t\right)/S_t$, in model (1). The second class of models utilizes a wider set of monthly predictors in model (1), corresponding to the sets of predictors previously used in the literature on monthly return regression models, as described in Table 1.[7] Examples include financial and macroeconomic variables that measure the state of the U.S. and global economy as well as oil-market specific variables that capture, for example, oil inventory dynamics, the degree of financialization, and hedging pressures in the oil futures market and in related futures markets. The construction of the data is described in the appendix. The third class of models relies on direct estimates of the risk premium from term structure models based on oil futures prices only (see Hamilton and Wu 2014).[8]

Despite the growing evidence that there is a statistically significant time-varying risk premium in the oil futures market, no one to date has quantified these risk premia in dollars, as required for the construction of measures of oil price expectations. The methodology proposed in section 3 allows us to derive consistent monthly estimates of the time-varying risk premium at each horizon of interest based on alternative specifications of equation (1). These estimates along with the corresponding estimate from the term structure model of Hamilton and Wu (2014) are reported in Figure 1. For expository purposes we focus on the 12-month horizon. Qualitatively similar results are obtained for all other horizons. Given our definition of $F_t^{12} = E_t\left(S_{t+12}\right) + RP_t^{12}$, a risk premium of -$5, for example, means that the implied oil price expectation exceeds the oil futures price by $5.

[7] Only two studies have used quarterly predictors (see Etula 2013, Acharya et al. 2013). These predictors are examined in section 8.

[8] Building on Ang and Piazzesi (2003), Hamilton and Wu propose a model of the time-varying risk premium that imposes an affine factor structure for oil futures prices. Estimation of the model only requires data for oil futures prices on the last day of each week. Time-varying risk premia are obtained as the difference between observed oil futures prices and the rational expectation of oil futures prices implied by the estimated term structure model.

Figure 1 allows a comparison of the risk premium estimates across our 17 model specifications. Figure 1 shows that all models agree that the risk premium prior to 2004 has been slightly negative on average. In 2004, however, there is a systematic downward shift in the average risk premium estimate, regardless of the model specification, accompanied by increased dispersion in the risk premium estimates across specifications and over time. The most striking insight provided by Figure 1 is that there is substantial disagreement on the magnitude and sign of the time-varying risk premium in the WTI futures market, especially in the last decade. Alternative estimates of the risk premium may differ by as much as $61 for the same month. The largest dispersion is observed for May 2008. The standard deviation of the risk premium estimates across alternative model specifications range from $0.36 in a given month to $14.47.

This point is important because the standard approach to recovering price expectations from futures prices is to subtract $\widehat{RP}_t^h$ from the futures price (e.g., Piazzesi and Swanson 2008). Figure 1 shows that this strategy would lead to a bewildering array of mutually inconsistent expectations estimates. Clearly, not all of the risk premium estimates in Figure 1 can be equally valid. This raises the question of which risk premium estimates we can rely on and which ought to be discarded. Answering this question is essential for constructing a unique and reliable measure of oil price expectations. In the next section we demonstrate how one can recover the most credible estimate of the risk premium and hence of the oil price expectations from any given set of candidate estimates.

5. How to select the most accurate estimate of the risk premium

Although the idea of measuring market expectations by risk adjusting futures prices is not new, the fact that different estimates of the risk premium imply very different estimates of the market expectation of the price creates a challenge for applied researchers interested in measuring the

price expectations that prevailed in futures markets. This situation is not unique to oil futures markets, but is a common feature of many futures markets.

Suppose that we are interested in constructing a measure of how the market expectation of the spot price, denoted by $E_t(S_{t+h})$, has evolved over time in the past. In the absence of a fully articulated economic decision problem, the conventional metric in assessing the accuracy of price expectations is their MSPE, defined as $E[S_{t+h} - F_t(S_{t+h})]^2$.[9] As noted earlier, standard arbitrage arguments imply that in the absence of a risk premium the futures price is the conditional expectation of the spot price, $F_t^h = E_t[S_{t+h}]$ (e.g., Pindyck 2001; Serletis 1991; Alquist and Kilian 2010). It is a well-known statistical result that the conditional expectation minimizes the MSPE under quadratic loss (e.g., Granger 1969; Granger and Newbold 1986). Hence, in the absence of a risk premium, F_t^h minimizes the MSPE. It follows immediately that in the presence of a time-varying risk premium, $F_t^h - RP_t^h$ instead attains the minimum MSPE.

While RP_t^h is not observable, it can be estimated based on the full sample, as described in section 3. Misspecifications of the risk premium model would be expected to reveal themselves based on a comparison of the MSPEs of F_t^h and $F_t^h - \widehat{RP}_t^h$. It follows immediately that any $\widehat{RP}_t^h$ that yields an expectation $F_t^h - \widehat{RP}_t^h$ with higher MSPE than F_t^h is inadmissible and should be discarded from further consideration. Moreover, it follows that the most credible risk premium estimate, among a given set of candidate estimates, is obtained by selecting the specification that delivers the largest MSPE reductions for the implied oil price expectation. Below we illustrate this general idea in the context of the problem of recovering expectations

[9] For example, Anderson et al. (2011) evaluate survey energy price expectations based on their MSPE. This approach is also standard in the literature on forecasting energy prices (e.g., Alquist et al. 2013).

about the future path of the spot price of crude oil from oil futures prices. The fact that the sets of predictors used in our payoff regressions were originally selected based on their ability to explain holding-period returns, as defined in footnote 1, rather than the payoff on oil futures contracts, helps address concerns about data mining biases.

5.1. Empirical results

Our measure of the predictive success of the expectation of the price of oil is the ratio of the MSPE of the risk-adjusted futures prices to the MSPE of the no-change forecast. This normalization is without loss of generality and facilitates the implementation of statistical tests for improvements in accuracy (see Alquist et al. 2013). We assess the statistical significance of the MSPE reductions based on the test of Clark and West (2007).

A useful starting point is Table 2 which focuses on basis regressions in the spirit of Fama and French's (1987, 1988) work on risk premia in non-oil commodity markets. The first column of results shows the MSPE ratios of the unadjusted oil futures price, which are generally slightly below one, indicating an improvement on the accuracy of the no-change forecast. The next column shows the corresponding MSPE ratios for the oil price expectation measure obtained by subtracting from the oil futures price the risk premium estimate obtained from regression (1) with x_t consisting only of the basis. Because for each horizon h the MSPE ratio of the risk-premium adjusted oil futures price exceeds that of the unadjusted oil futures price, we conclude that this specification is inadmissible because it has a strictly higher MSPE than the unadjusted oil futures price. In other words, there is no evidence of a time-varying risk premium. The last column shows complementary evidence based on model (1) with $\beta_h = 0 \; \forall \; h$. This model only allows for a constant risk premium. Again at each horizon the MSPE ratio exceeds that of the

benchmark in the first column, so this model as well is inadmissible.[10]

Results for model (1) with the extended set of predictors and for the term structure model of Hamilton and Wu (2014) are reported in Table 3. The latter model is labelled HW. The other model specifications are labelled as shown in Table 1. There is little evidence that the regressors that proved successful in predicting future returns are also successful at predicting the payoff. Table 3 shows that, after adjusting for the estimated risk premium, the MSPE ratios decline systematically across all horizons in only one of sixteen cases. The risk-adjusted futures price based on the HW model yields reductions in the MSPE ratio by an additional 19 to 36 percentage points compared with the futures price, depending on the horizon. Thus, by the model selection criterion introduced in this section, only the HW specification should be used in estimating the market expectation of the price of oil.

5.2. An encompassing approach to predicting the risk premium

One concern with the choice of predictors in the extant literature on the time-varying risk premium is that there are many specifications that are considered successful at capturing the time-varying risk premium, but no two specifications agree on the relevant set of predictors. This observation suggests that there is room for improving on the accuracy of existing specifications of x_t by considering all predictors jointly rather than focusing on subsets. We therefore also considered a generalized payoff regression model that encompasses all 30 individual predictors utilized so far, excluding only the predictors in the BS model, which are not available for the full sample. This all-predictor regression approach allows for the fact that the predictors may be

[10] It is worth emphasizing that we are not concerned with testing the null hypothesis of the absence of a risk premium. It can be shown that $H_0 : \beta_h = 0$ in model (1) cannot not be rejected at any horizon in favor of a time-varying risk premium. This result is not informative, however, because adjusting the futures price for the estimated risk premium may still improve the predictive accuracy of oil futures prices, even if the null is not rejected. The latter question can only be addressed as shown in Table 2.

mutually correlated. We are not interested in the predictive power of individual regressors at this point so much as in their joint predictive power. The second column of Table 4 suggests that there are substantial gains in accuracy from jointly considering a larger set of predictors. The expectations measure implied by the all-predictor model has much lower MSPE ratios at all horizons than any other individual model with the exception of the HW model. Although the HW model remains more accurate at horizons 6, 9, and 12, it is less accurate at horizon 3.[11]

A possible concern with this model is that the larger number of predictors in the payoff regression may invite overfitting in small samples. The fact that this model includes 30 predictors (rather than a maximum of seven predictors as in the individual payoff regression specifications) would be expected to improve the fit of the model in small samples, even if the additional predictors were irrelevant. A simulation exercise reveals that including 30 redundant white noise predictors on average lowers the MSPE ratio at the 12-month horizon by about 0.07, suggesting that the MSPE ratio of 0.57 in Table 4 should be adjusted to 0.64, at which point the all-predictor model is clearly dominated by the HW model. Further simulation analysis shows that allowing some of the 30 redundant predictors to be persistent may increase the bias in the MSPE ratio to 0.38, suggesting that the true MSPE ratio of the all-predictor model may only be 0.95. In contrast, when using only seven predictors, as in the B2 specification, for example, the overfitting bias is negligible. Although treating all predictors as invalid is an extreme assumption, this thought experiment suggests caution in interpreting the results for the all-predictor model.

A natural question is whether this overfitting problem may be mitigated by selecting a subset of the 30 predictors based on pre-tests. Such a pre-test strategy closely mirrors the

[11] Shortening the evaluation period and including the BS predictor improves the accuracy of the all-predictor model at horizons 3, 9 and 12, clearly outperforming the HW model at horizon 3 and marginally so at horizon 12.

approach taken in the finance literature, which often judges the usefulness of a predictor based on its statistical significance. Table 5 shows that of the 30 predictors only the cross-market hedging pressure from the silver futures market appears statistically significant at all horizons based on two-sided t-tests. Among the other predictors, unexpected changes in U.S. industrial production, the composite leading indicator of global economic growth, and the basis in oil futures markets are significant at all but one horizon, while the CRSP equity index, the change in the U.S. default premium, the junk bond premium, and the growth rate of dollar open interest are statistically significant at two horizons. This evidence suggests that we consider a second payoff regression based on the subset of predictors that are statistically significant at the 10% level in the unrestricted regression. At the 3-month horizon, for example, the pre-test model only includes 7 predictors and at the 12-month horizon only 12 predictors compared with the original 30 predictors, greatly reducing concerns about potential overfitting. Although the results after pre-testing remain more accurate than any of the individual payoff regressions in Tables 2 and 3, the third column in Table 4 shows that dropping statistically insignificant predictors systematically lowers the accuracy of the expectations measure at all horizons compared with the unrestricted model in the second column. The accuracy of the after pre-testing model at all horizons is dominated by that of the HW model.

An alternative approach to combining information from many predictors without increasing the dimensionality of the original payoff regression models is model averaging. The fourth column of Table 4 shows results obtained by equal-weighted averaging of the oil price expectations obtained by the 17 payoff-regression models of the time-varying risk premium considered in Tables 2 and 3. The latter specification is less accurate than the unadjusted oil futures price except at horizon 3, and the gain in accuracy at horizon 3 is negligible. The average

oil price expectation also is less accurate than the HW model. This evidence reinforces our earlier concerns with the all-predictor model.

5.3. Other criteria for evaluating estimates of oil price expectations

There are additional criteria for judging the accuracy of alternative oil price expectation measures, however, that can help in deciding between specifications of similarly high statistical accuracy. For example, one would not expect longer-term oil price expectations to be highly volatile. Except during times of major events affecting the market for oil, they should evolve smoothly over time. In fact, one would expect longer-horizon oil price expectations to evolve more smoothly than the underlying oil futures price. Figures 2 and 3 allow us to address this question.

Figure 2 shows the risk-adjusted futures price based on the all-predictor model for horizons $h \in \{3,6,9,12\}$. It also reports the unadjusted F_t^h for comparison. The gap between these lines is the risk premium. Our discussion focuses on the 12-month horizon to conserve space. In the early 1990s, there is little difference between the oil futures price and the risk-premium adjusted oil price futures price. Between late 1997 and June 2005, however, the risk-premium adjusted WTI futures price exceeds the WTI oil futures price by $7 on average, indicating that the market expected the price of oil to be higher than indicated by the oil futures price. Starting in mid-2005, the risk-premium adjusted oil futures price becomes increasingly unstable. For example, the 12-month-ahead oil price expectation reaches $104 in October 2006, but drops as low as $64 in November 2007, only to rise again to $111 by May 2008. Likewise, between June 2009 and June 2013, when oil markets had stabilized again following the global financial crisis, the 12-month-ahead oil price expectation in Figure 2 fluctuates between $77 and $133. In fact, the risk-premium adjusted oil futures price is systematically more volatile than the

oil futures price. It seems difficult to reconcile this erratic pattern with the perception that longer-term oil price expectations should be smoother than the observed oil futures price. Moreover, oil price expectations reach their all-time high of $133 in April 2011, long after the peak in the spot price. Likewise, the second highest oil price expectation of $124 is recorded in March 2012, long after the historical peak in the spot and futures price of oil in mid-2008. It is not clear what could account for such a pattern.[12]

In contrast, the evolution of the risk-premium adjusted futures price based on the HW model in Figure 3 is consistently smoother than that of the unadjusted oil futures price. Between January 2003 and January 2008, during the surge in the spot price of oil, the time-varying risk premium on average was -$12, although there are months when it reached -$28. During this time period, the risk-premium adjusted oil futures price consistently exceeded the oil futures price. Between February 2008 and September 2008 the 12-month risk premium rose sharply to $19 on average, reaching a peak of $37. In other words, the risk-premium adjusted oil futures price during this period was much lower than the oil futures price. For the remainder of the sample it fluctuates between -$23 and +$17 with an average of -$2. We conclude that the oil price expectation implied by the HW risk premium model not only is more accurate in predicting S_{t+h} at longer horizons, but also is economically more plausible than the measure of oil price expectations generated by the all-predictor model.

6. A retrospective analysis of the market's oil price expectations

The time series of oil price expectations in Figure 3 provides a superior alternative to the existing measures of oil price expectations used for economic modelling including oil futures prices and no-change predictions. Closer examination of the lower right panel of Figure 3 reveals several

[12] Qualitatively similar results are obtained if we shorten the sample and include the BS specification in the all-predictor models.

intriguing results. For example, while there is no evidence that the market anticipated the collapse of oil prices in the second half of 2008, even when the spot price unexpectedly reached $134 in June of 2008, market participants did not expect the price of oil to remain at that level for another year. An equally interesting finding is that from 2010 until mid-2013 the 12-month-ahead oil price expectation stabilized at a level near $90, notwithstanding the short-term fluctuations in the oil futures price about that level. Finally, Figure 3 also indicates that the 12-month-ahead oil price expectations increased with few exceptions throughout the period from early 2003 to mid-2008, from $30 initially to a peak of $100 in June 2008. This result is in marked contrast to the term structure of oil futures during this period, which remained largely flat or even downward sloping, resulting in persistent prediction errors. Our analysis implies that the poor predictive accuracy of the unadjusted oil futures prices between 2003 and mid-2008 reflected not so much systematic errors in expectations, but the omission of the time-varying risk premium.

To make this point more clearly, in Figure 4 we focus on four specific points in time, each selected to be 15 months apart from the previous observation (April 2003, July 2004, October 2005, January 2007). Figure 4 illustrates that accounting for the HW risk premium typically helps close the persistent gap between oil price expectations and realizations across the term structure of futures prices, although there still are some points in time when substantial prediction errors occur even after accounting for the risk premium. Because nothing in the construction of the HW model ensures that adjusting for the risk premium would explain the poor predictive accuracy of oil futures prices during this period, this pattern further increases our confidence in the ability of the HW model to capture market expectations. Being able to identify predictive failures of the oil futures prices driven by variation in the risk premium, as in this

example, is important, more generally, because reliable measures of oil price expectations are a precondition for evaluating the empirical content of models of learning and of peso problems in futures markets (e.g., Timmermann 1993).

7. Implications for real-time oil price forecasts

Our analysis in section 5 provided compelling evidence that oil futures prices should not be used as oil price expectations in estimating microeconomic models of the purchases of energy-using durables. The use of oil futures prices as oil price expectations is not restricted to microeconomic applications, however. Oil futures prices have also been used as out-of-sample forecasts of the price of oil in macroeconomics. This approach is very common at central banks and international organizations, for example. The popularity of this approach has several reasons. First, futures prices are simple to use and readily available in real time. Second, there is a reluctance to depart from what is viewed as the collective wisdom of the financial market which presumably knows better than any individual oil price forecaster. Relying on what is perceived to be the market expectation also absolves the forecaster from any culpability for forecast errors because no one can reasonably be expected to beat the market. Third, there is evidence that futures prices have outperformed other oil price forecasts on average at least at some forecast horizons, although their predictive accuracy has varied substantially over time. Moreover, until recently there were few alternatives available to oil price forecasters (e.g., Alquist, Kilian and Vigfusson 2013). Fourth, while it is well understood that time-varying risk premia would invalidate the use of oil futures prices as oil price forecasts, it has proved difficult to reject the absence of a time-varying risk premium based on the traditional statistical tests of forecast efficiency proposed by Fama and French (1987, 1988) (e.g., Alquist and Kilian 2010).

Although our approach to estimating historical oil price expectations was not designed to

generate out-of-sample forecasts of the price of oil, one may nevertheless wonder how the HW model, which performed best in our original analysis, would perform at the latter task compared with the unadjusted oil futures price. We address this question below. It is well known that - even if the model is correct – predictive success based on the full sample need not generalize to out-of-sample forecasting, given the bias-variance trade-off in estimating forecasting models (see Kilian 2015). Producing risk-adjusted out-of-sample forecasts involves the recursive estimation of the time-varying risk premium using only data available to the forecaster at the time the forecast is generated. The latter real-time data constraint does not affect the HW model, which only requires data on oil futures prices that are never revised and that are immediately available.

By construction, the benchmark results for the unadjusted futures price, as shown in the first column of Table 6, are unchanged when doing out-of-sample analysis. The next column of Table 6 reports results based on recursive real-time estimates of the HW model with one important difference. Whereas the original specification of the HW model imposed a structural break in the term structure model in January 2015, the recursively estimated term structure model does not split the sample in January 2005 on the grounds that such a break would not have been detected in real time. The implied risk-premium adjusted real-time forecasts have higher MSPE than the unadjusted futures price at all horizons. There is no improvement on the use of oil futures prices as out-of-sample forecasts.

There are three potential reasons for this disappointing result. First, the structure of the HW model implies that forecasts are generated at the end of the third week of the current month, whereas the convention in out-of-sample forecasting is that we forecast using all of the information available at the end of the current month. The third column of Table 6 attempts to remedy this problem in an ad hoc fashion by adjusting the level of the HW oil price forecast by

the change in the daily futures price of maturity h between the day on which the HW forecast was generated and the last trading day of the month. Table 6 shows that this timing correction improves the accuracy of the HW forecast, but not by much. The MSPE ratios remain above one.

A second potential explanation is that recursive real-time forecasts are based on much smaller samples than the full-sample analysis in Table 3, resulting in less accurately estimated model parameters and hence less accurate forecasts. We quantify this effect in the fourth column of Table 6, which shows results for the HW model evaluated at the full-sample parameters estimates (but without imposing the structural break). Table 6 shows that estimation uncertainty must be important because these MSPE ratios are below one at all horizons and statistically significant at horizons 9 and 12. They are in fact only slightly less accurate than the full-sample version of the HW model without break. This evidence helps us understand the inferior performance in the second column, but offers no practical solution for out-of-sample forecasters.

There is, however, a third explanation, namely that the low out-of-sample accuracy reflects the fact that in the out-of-sample analysis we ignored the structural break in 2005 that Hamilton and Wu (2004) allowed for. The importance of breaking the sample in January 2005 is reinforced by the evidence in Figure 1 which shows that all payoff-regression models of the time-varying risk premium generate a mean shift in the time-varying risk premium around this time. Given that the HW model does not involve additional predictors, it cannot generate this shift endogenously. Clearly, however, a forecaster would not have known about this shift in real time, so not imposing this break in the estimation of the term structure model is only fair.

Some researchers favor the use of rolling estimation windows in the presence of unknown structural breaks. This approach does not work here. Estimating the time-varying risk premium based on a rolling window, as shown in the fifth column of Table 6, yields MSPE ratios

very similar to the recursive forecasts in the second column. An alternative and more promising solution is simply to discard the earlier data and base the estimation of the model on post-break data. This post-break model is evaluated in the sixth column of Table 6. We start the estimation sample in January 2005 and allow for 48 observations in the initial recursive sample. Table 6 indicates that, notwithstanding the short estimation samples, the post-break HW model performs much better with MSPE ratios well below one and highly statistically significant at all horizons.

The seventh column of Table 6 applies the timing adjustment of column 3 to the post-break model, further improving the out-of-sample accuracy of the HW model. Even granting that these results are based on an evaluation period much too short to draw firm conclusions, these results suggest that simple adjustments can go a long way toward restoring the predictive accuracy of the HW model even out of sample. In fact, controlling for the evaluation period, the last HW specification has lower MSPE ratios than the unadjusted futures price at horizons 3, 6 and 9. Only at horizon 12, the oil futures price remains more accurate. We conclude that this new approach to out-of-sample forecasting deserves further study, as more data become available, and that it ought to be compared systematically to alternative out-of-sample forecasting approaches (e.g., Baumeister and Kilian 2015).[13]

8. Sensitivity analysis

Although most empirical studies of the time-varying risk premium in the oil futures market rely on monthly predictors, the studies by Acharya et al. (2013) and Etula (2013) are based on quarterly data. As the appendix shows, several of the quarterly model specifications drawn from

[13] For completeness, in the appendix, we also investigated the out-of-sample accuracy of the other 16 models of the time-varying risk premium whose predictors are not necessarily available in real time. Even when using fully revised data, however, these models are systematically less accurate than the oil futures price. This conclusion also applies to the models of Pagano and Pisani (2009), which were originally designed for out-of-sample forecasting. The all-predictor model, to the extent that there are enough degrees of freedom to estimate this model at all in real time, is even less accurate in out-of-sample than the other payoff regressions, given the much larger number of model parameters to be estimated.

the literature are admissible in that the implied oil price expectations are more accurate than the oil futures price. As in the monthly case, however, the all-predictor model substantially improves on the accuracy of all the original specifications, except the HW model. The HW model is more accurate than the all-predictor model at horizons 2 and 3 and about as accurate at horizon 4. At horizon 1, the all-predictor model is clearly more accurate. As in the monthly setting, however, there are concerns that the all-predictor model may be overfitting, and only the HW model provides economically plausible estimates of the oil price expectations. The evolution of the quarterly HW oil price expectations is similar to that of the monthly estimates. Finally, none of the original quarterly models improve on the real-time accuracy of the quarterly oil price forecasts based on oil futures prices, but the out-of-sample accuracy of the HW model may be improved by using a specification analogous to that in Table 6 (column 7).[14]

9. Concluding remarks

A time-varying risk premium renders the futures price a poor measure of the market expectation of the price of the underlying asset. Estimates of the risk premium may be used to construct improved measures of this expectation. Typically, however, these estimates differ greatly across model specifications, creating uncertainty about the magnitude of the implied market expectation of the price of oil. We exploited the theoretical insight that the risk-premium adjusted futures price is the conditional expectation of the price of oil (and hence minimizes the MSPE) to rank alternative risk premium estimates based on their MSPE.

We illustrated this approach by quantifying the market's expectation of the price of oil.

[14] The latter HW specification has lower out-of-sample MSPEs over the last six years of our sample than the oil futures price at horizons of 1, 2 and 3 quarters and similar accuracy 4 quarters ahead. It also has much higher real-time accuracy than the no-change forecast of the spot price at all four horizons.

Our preferred model of the time-varying risk premium was a term structure model of the oil futures market developed by Hamilton and Wu (2014). The expectations measure implied by this model had higher accuracy than other specifications and was economically more plausible. We highlighted that this new measure of the market expectation of the spot price of oil is useful in modelling expectations across a wide range of economic models dealing with energy prices. We also demonstrated how correcting for the time-varying risk premium helps explain the persistent errors implied by the use of futures prices as oil price expectations between 2003 and mid-2008.

In contrast, alternative risk premium specifications motivated by the existing literature tend to generate inaccurate estimates of oil price expectations. We demonstrated that by combining all the predictors used in this literature in the same regression model of the time-varying risk premium, the accuracy of the oil price expectation can be substantially improved, but at the cost of potential overfitting. Moreover, the implied oil price expectations remain less economically plausible and often less accurate than those based on the Hamilton-Wu model.

In extracting the market expectation of the price of oil, it is essential to estimate the risk premium based on the full sample. This approach provides the most efficient estimate of the oil price expected by the market at each point in time in the past, which is the relevant expectations measure, for example, in estimating models of purchases of automobiles as in Kahn (1986), Busse, Knittel and Zettelmeyer (2013) or Allcott and Wozny (2014). In contrast, if the objective is to improve the accuracy of out-of-sample forecasts of the price of oil, as is common at central banks, real-time estimates of the risk premium are required. We provided additional tentative evidence that recursive out-of-sample forecasts based on the Hamilton-Wu model with suitable adjustments are able to improve on the accuracy of the unadjusted oil futures price at horizons up to nine months.

One advantage of our approach is that the information set used by financial market participants is wider than that embodied in conventional regression-based measures of oil price expectations. It embodies, for example, the market's assessment of the political stability of the Middle East, of the prospects for new oil discoveries, or of the profitability of new extraction technologies, which is difficult to capture by conventional methods. It should be noted that this oil price expectations measure may also be used for constructing new measures of oil price shocks, as illustrated in Baumeister and Kilian (2016a). Understanding the magnitude of the price shocks experienced by participants in financial markets is obviously important in modelling the transmission of these shocks and in designing policy responses.

While we chose to illustrate our procedure for recovering the market expectation of the price in the context of the oil futures market, it is much more widely applicable. The same methodology may be applied to futures prices for foreign exchange, interest rates or other commodities, for example, whenever there is disagreement between alternative models of the time-varying risk premium.

References

Acharya, V., L. Lochstoer, and T. Ramadorai (2013), "Limits to Arbitrage and Hedging: Evidence from Commodity Markets," *Journal of Financial Economics*, 109, 441-465.

Allcott, H., and N. Wozny (2014), "Gasoline Prices, Fuel Economy, and the Energy Paradox," *Review of Economics and Statistics*, 96, 779-795.

Alquist, R., and Kilian, L. (2010), "What Do We Learn from the Price of Crude Oil Futures?" *Journal of Applied Econometrics*, 25, 539-573.

Alquist, R., Kilian, L., and R.J. Vigfusson (2013), "Forecasting the Price of Oil," in: Elliott, G., and A. Timmermann (eds.), *Handbook of Economic Forecasting*, 2, Amsterdam: North-Holland, 427-507.

Anderson, S.T., Kellogg, R., Sallee, J.M., and R.T. Curtin (2011), "Forecasting Gasoline Prices Using Consumer Surveys," *American Economic Review: Papers & Proceedings*, 101, 110-114.

Ang, A., and M. Piazzesi (2003), "A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables," *Journal of Monetary Economics*, 50, 745-787.

Baumeister, C., and L. Kilian (2015), "Forecasting the Real Price of Oil in a Changing World: A Forecast Combination Approach," *Journal of Business and Economic Statistics*, 33, 338-351.

Baumeister, C., and L. Kilian (2016a), "Forty Years of Oil Price Fluctuations: Why the Price of Oil May Still Surprise Us," *Journal of Economic Perspectives*, 30, 139-160.

Baumeister, C., and L. Kilian (2016b), "Understanding the Decline in the Price of Oil since June 2014," *Journal of the Association of Environmental and Resource Economists*, 3, 131-158.

Bessembinder, H. (1992), "Systematic Risk, Hedging Pressure, and Risk Premiums in Futures Markets," *Review of Financial Studies*, 5, 637-667.

Bessembinder, H., and K. Chan (1992), "Time-Varying Risk Premia and Forecastable Returns in Futures Markets," *Journal of Financial Economics*, 32, 169-193.

Bessembinder, H., and P. Seguin (1993), "Price Volatility, Trading Volume and Market Depth: Evidence from Futures Markets," *Journal of Financial and Quantitative Analysis*, 28, 21-39.

Busse, M., Knittel, C.R., and F. Zettelmeyer (2013), "Are Consumers Myopic? Evidence from New and Used Car Purchases," *American Economic Review*, 103, 220-256.

Clark, T.E., and K.D. West (2007), "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models," *Journal of Econometrics*, 138, 291-311.

De Roon, F.A., T.E. Nijman, and C. Veld (2000), "Hedging Pressure Effects in Futures Markets," *Journal of Finance*, 55, 1437-1456.

Diebold, F.X., and R.S. Mariano (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Etula, E. (2013), "Broker-Dealer Risk Appetite and Commodity Returns," *Journal of Financial Econometrics*, 11, 486-521.

Fama, E.F., and K.R. French (1987), "Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage," *Journal of Business*, 60, 55-73.

Fama, E.F., and K.R. French (1988), "Business Cycles and the Behavior of Metals Prices," *Journal of Finance*, 43, 1075-1093.

Gorton, G., F. Hayashi, and G. Rouwenhorst (2013), "The Fundamentals of Commodity Futures Returns," *Review of Finance*, 17, 35-105.

Granger, C.W.J. (1969), "Prediction with a Generalized Cost of Error Function," *Operations Research Quarterly*, 20, 199-207.

Granger, C.W.J., and P. Newbold (1986), *Forecasting Economic Time Series*, 2nd ed., Academic Press: San Diego.

Hamilton, J.D., and J.C. Wu (2014), "Risk Premia in Crude Oil Futures Prices," *Journal of International Money and Finance*, 42, 9-37.

Hong, H., and M. Yogo (2012), "What Does Futures Market Interest Tell Us about the Macroeconomy and Asset Prices?" *Journal of Financial Economics*, 105, 473-490.

Kahn, J.A. (1986), "Gasoline Prices and the Used Automobile Market: A Rational Expectations Asset Price Approach," *Quarterly Journal of Economics*, 101, 323-339.

Kellogg, R. (2014), "The Effect of Uncertainty on Investment: Evidence from Texas Oil Drilling," *American Economic Review*, 104, 1698-1734.

Kilian, L. (2015), "Comment on Francis X. Diebold's 'Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests'," *Journal of Business and Economic Statistics*, 33, 13-17.

Pagano, P., and M. Pisani (2009), "Risk Adjusted Forecasts of Oil Prices," *B.E. Journal of Macroeconomics*, 9, article 24, 1-25.

Piazzesi, M., and E.T. Swanson (2008), "Futures Prices as Risk-Adjusted Forecasts of Monetary Policy," *Journal of Monetary Economics*, 55, 677-691.

Pindyck, R.S. (2001), "The Dynamics of Commodity Spot and Futures Markets: A Primer," *Energy Journal*, 22, 1-29.

Sadorsky, P. (2002), "Time-Varying Risk Premiums in Petroleum Futures Prices," *Energy Economics*, 24, 539-556.

Serletis, A. (1991), "Rational Expectations, Risk and Efficiency in Energy Futures Markets," *Energy Economics*, 13, 111-115.

Singleton, K.J. (2014), "Investor Flows and the 2008 Boom/Bust in Oil Prices," *Management Science*, 60, 300-318.

Timmermann, A. (1993), "How Learning in Financial Markets Generates Excess Volatility and Predictability in Stock Prices," *Quarterly Journal of Economics*, 108, 1135-1145.

Table 1. Monthly Predictor Variables for Payoff on Oil Futures Contracts

Article	Model	Predictors
Bessembinder (1992)	B1	Returns on CRSP value-weighted equity index
	B2	Returns on CRSP value-weighted equity index
		Unexpected CPI inflation
		Change in expected CPI inflation
		Change in 3-month T-bill rate
		Change in the term structure (20YGB – 3-month T-bill)
		Change in default premium (Baa – 20YGB)
		Unexpected change in U.S. industrial production
Bessembinder and Chan (1992)	BC	Dividend yield on CRSP value-weighted equity index
		3-month T-bill rate
		Junk bond premium (Baa – Aaa)
Sadorsky (2002)	S	Return on dividend yield on S&P 500 common stock portfolio
		Return on junk bond premium (Baa – Aaa)
		Return on 3-month T-bill rate
		Market portfolio excess return
De Roon, Nijman, and Veld (2000)	DNV1	Returns on S&P 500 stock price index
		Own-market hedging pressure
		Cross-market hedging pressure for gold, silver, platinum, heating oil
	DNV2	Own-market hedging pressure scaled by its standard deviation
		Own-market price pressure scaled by its standard deviation
	DNV3	DNV1 + own-market price pressure
Gorton, Hayashi, and Rouwenhorst (2013)	GHR1	Normalized U.S. crude oil commercial inventories (no SPR)
	GHR2	Own-market hedging pressure
Hong and Yogo (2012)	HY1	1-month T-bill rate
		Yield spread (Aaa – 1MTbill)
		Basis by horizon
	HY2	HY1 + growth rate of oil market dollar open interest
Pagano and Pisani (2009)	PP1	Degree of capacity utilization in U.S. manufacturing
	PP2	Term spreads (2YGB–1YGB, 5YGB–2YGB, 10YGB–5YGB)
	PP3	Composite leading indicator for OECD + 6 NMEs
Bessembinder and Seguin (1993)	BS	Ratio of trading volume of oil futures contracts to open interest by horizon

NOTES: The sample period is 1986.1-2014.6 except for the series from the CRSP database which are only available until 2013.12 and the series in BS which start only in 1989.9 for horizons 3, 6, and 9, and in 1992.4 for horizon 12.

Table 2. Predictive Accuracy of Risk-Adjusted Oil Futures Prices Based on Full-Sample Estimates of the Risk Premium

MSPE Ratio for Evaluation Period: 1992.1-2014.6

Horizon h	No Risk Premium F_t^h	Time-Varying Risk Premium $F_t^h\left(1+\hat{\alpha}+\hat{\beta}\left(F_t^h-S_t\right)/S_t\right)$	Constant Risk Premium $F_t^h\left(1+\hat{\alpha}\right)$
3	**0.987**	1.035	1.035
6	**0.982**	1.073	1.082
9	**0.949**	1.074	1.087
12	**0.882***	1.041	1.043

NOTES: All MSPE ratios have been normalized relative to the monthly no-change forecast. Boldface indicates an improvement on the monthly no-change forecast. Statistically significant improvements test are marked using * (5% significance level) and ** (10% significance level). The underlying risk-premium estimates are based on the full sample. MSPE reductions are evaluated based on the tests of Diebold and Mariano (1995) and Clark-West (2007), as appropriate.

Table 3. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Full-Sample Estimates of the Risk Premium

MSPE Ratio for Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	B1	B2	BC	S	DNV1	DNV2	DNV3
3	**0.987**	**0.972***	**0.880****	1.022	**0.992***	**0.927***	1.043	**0.927***
6	**0.982**	1.054	**0.964***	1.073	1.063	1.005	1.095	1.005
9	**0.949**	1.063	1.002	1.078	1.068	1.040	1.122	1.041
12	**0.882***	1.013	**0.901***	1.044	1.004	**0.923***	1.082	**0.923***

Horizon h	GHR1	GHR2	HY1	HY2	PP1	PP2	PP3	BS	HW
3	**0.991***	1.044	1.010	1.007	1.027	1.013	**0.964***	1.027	**0.794***
6	1.015	1.102	1.046	**0.988****	1.089	1.053	1.051	1.073	**0.667***
9	**0.997***	1.118	1.075	**0.986****	1.117	1.038	1.080	1.080	**0.592***
12	**0.831***	1.088	1.071	1.022	1.084	**0.987**	1.045	1.045	**0.535***

NOTES: See Table 2. The evaluation period for B1, B2, BC and S for h=3 ends in 2013.12 and that for BS at h=12 starts only in 1992.4 due to data constraints.

Table 4. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Full-Sample Estimates of the Risk Premium
MSPE Ratio for Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	All predictors	After pre-testing	Model Averaging	HW
3	0.987	0.711*	0.796*	0.976*	0.794*
6	0.982	0.738*	0.885*	1.035	0.667*
9	0.949	0.764*	0.862*	1.045	0.592*
12	0.882*	0.568*	0.667*	0.980**	0.535*

NOTES: See Table 2. The set of all predictors includes all variables in Table 3 except for market liquidity because the 12-month horizon liquidity series only starts in 1992.4. The pre-test consists of a two-sided t-test at the 10% level. The results of the pre-test are reported in Table 5. The estimates for $h = 3$ end in 2013.12, reflecting data constraints for some predictors. The model-averaging results refer to the equal-weighted average of the oil price expectations implied by the 17 payoff regression models of the time-varying risk premium in Tables 2 and 3 (or based on 16 models when the BS specification is infeasible).

Table 5. *p*-values for *t*-Tests on the Predictors in the Unrestricted Payoff Regression Model

		Horizon (Months)			
		3	6	9	12
B2	Returns on CRSP value-weighted equity index	0.124	**0.079**	**0.085**	0.108
	Change in 3-month T-bill rate	0.344	0.870	0.109	**0.008**
	Change in the term structure	0.330	0.942	0.822	0.974
	Change in default premium	**0.078**	**0.154**	0.428	0.289
	Unexpected change in U.S. industrial production	**0.055**	**0.049**	**0.021**	0.101
	Change in expected CPI inflation	0.617	0.287	0.826	**0.100**
	Unexpected CPI inflation	0.944	0.700	0.863	0.530
DNV1	Returns on S&P 500 stock price index	0.938	0.898	0.197	**0.025**
	Gold hedging pressure	**0.005**	0.130	0.667	0.235
	Silver hedging pressure	**0.095**	**0.000**	**0.001**	**0.000**
	Platinum hedging pressure	0.189	0.887	0.170	**0.050**
	Heating oil hedging pressure	0.923	0.307	0.705	0.786
	Own-market hedging pressure	0.603	0.912	0.996	**0.080**
GHR1	Normalized U.S. crude oil commercial inventories (no SPR)	**0.051**	0.271	0.149	**0.009**
HY1	1-month T-bill rate	0.830	0.623	0.634	0.589
	Yield spread	0.522	0.966	0.474	0.492
	Growth rate of oil market dollar open interest	0.374	**0.011**	**0.026**	0.375
	Basis	0.126	**0.046**	**0.004**	**0.015**
BC	Dividend yield on CRSP value-weighted equity index	0.879	0.993	0.745	0.889
	3-month T-bill rate	0.866	0.492	0.837	0.840
	Junk bond premium	**0.003**	**0.077**	0.148	0.423
S	Return on dividend yield on S&P 500 common stock portfolio	0.977	0.881	0.173	**0.028**
	Return on junk bond premium	0.727	0.588	0.814	0.189
	Market portfolio excess return	0.877	0.504	0.454	0.628
PP1	Degree of capacity utilization in U.S. manufacturing	0.895	0.835	0.661	0.566
PP2	Term spreads 1	0.943	0.995	0.887	**0.015**
	Term spreads 2	0.258	0.621	0.913	**0.029**
	Term spreads 3	0.424	0.836	0.856	**0.065**
PP3	Composite leading indicator for OECD + 6 NMEs	**0.004**	**0.003**	**0.082**	0.811
DNV3	Own-market price pressure	0.919	0.811	0.308	0.422

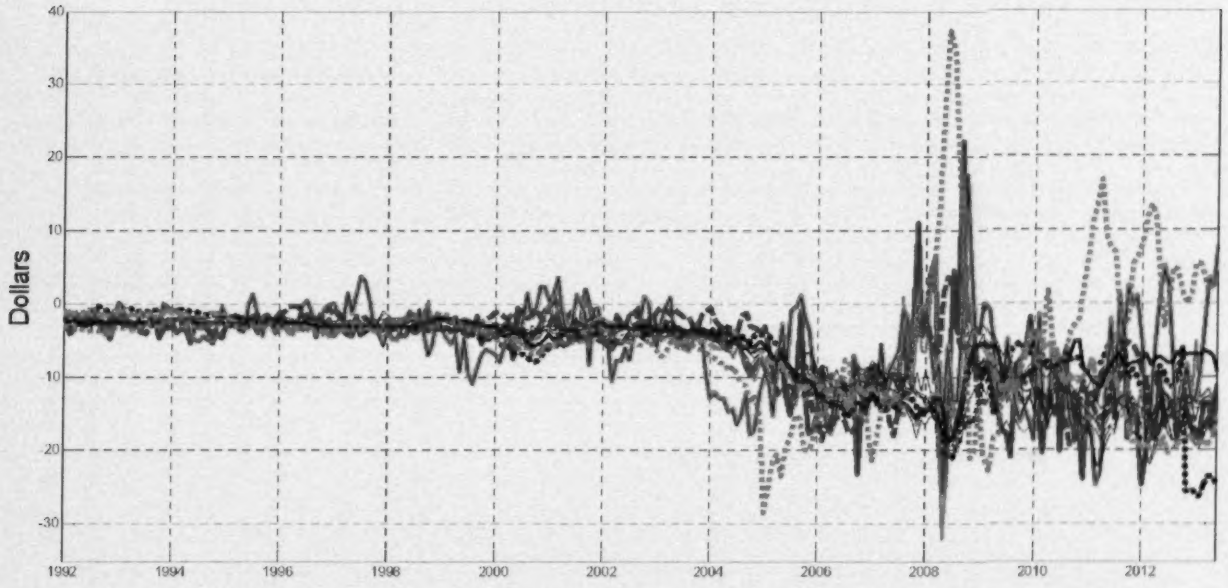
NOTES: The unrestricted payoff regression includes all 30 predictors included in the models in Table 4 except for BS. Boldface indicates statistical significance of the two-sided t-test of the null of no predictability at the 10% level. Only the statistically significant predictors are retained in the payoff regression labeled "After Pre-testing" in Table 4.

Table 6. MSPE Ratio of Risk-Adjusted Out-of-Sample Forecasts of the Spot Price Based on HW Term-Structure Model

	Evaluation period							
		1992.1:2014.6 Recursive Window			Rolling Window (60 months)	2009.1-2014.6 Recursive Window		
Horizon h	F_t^h	Baseline	Alternative 1	Alternative 2	Baseline	Post-break	Alternative 1: Post-break	F_t^h
3	**0.987**	1.083	1.001	**0.987**	1.160	**0.871***	**0.750***	**0.853***
6	**0.982**	1.206	1.158	**0.681**	1.242	**0.676***	**0.628***	**0.743***
9	**0.949**	1.365	1.318	**0.601****	1.275	**0.596***	**0.539***	**0.628***
12	**0.882***	1.511	1.481	**0.539***	1.227	**0.629***	**0.584***	**0.549***

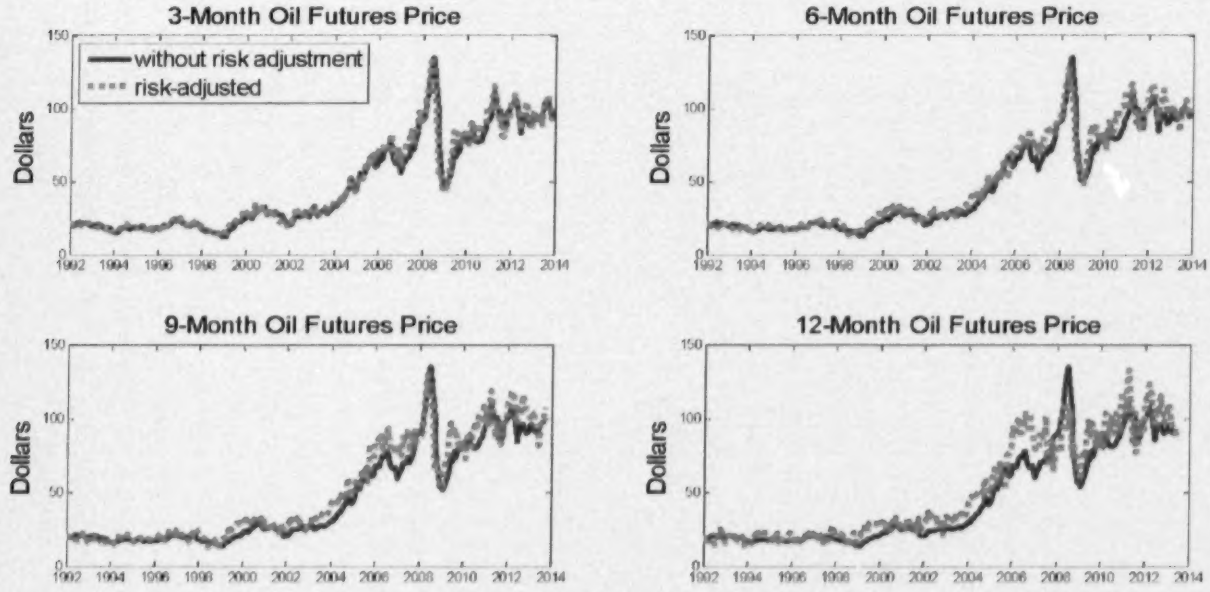
NOTES: Boldface indicates improvements on the no-change forecast. Statistically significant improvements test are marked using * (5% significance level) and ** (10% significance level). The baseline model corresponds to the HW model in Tables 3 and 4, except that the model is estimated recursively without allowing for a break in 2005.1. Alternative 1 refers to the same model, except we add to the HW forecast the change in the daily oil futures price of maturity h between the day on which the forecast is generated by the HW model (in the middle of the current month) and the last trading day of that month. Alternative 2 refers to the recursively evaluated HW model without breaks with the full-sample parameter estimates of the same model imposed. The post-break HW model estimates are based on data starting in 2005.1 (after the break date imposed in the original HW full-sample specification) and compared to the corresponding futures price benchmark. The initial recursive model is estimated based on 48 monthly observations. We follow the literature in assessing the statistical significance of the recursive MSPE reductions based on the test of Clark and West (2007). This test (like similar tests in the literature) is biased toward rejecting the null of equal MSPEs because it tests the null of no predictability in population rather than the null of equal out-of-sample MSPEs (see Inoue and Kilian 2004). Thus, these test results have to be interpreted with caution. For further discussion of the problem of out-of-sample inference see Kilian (2015).

Figure 1: Alternative Monthly Estimates of the Time-Varying Risk Premium in the Oil Futures Market at the 12-Month Horizon



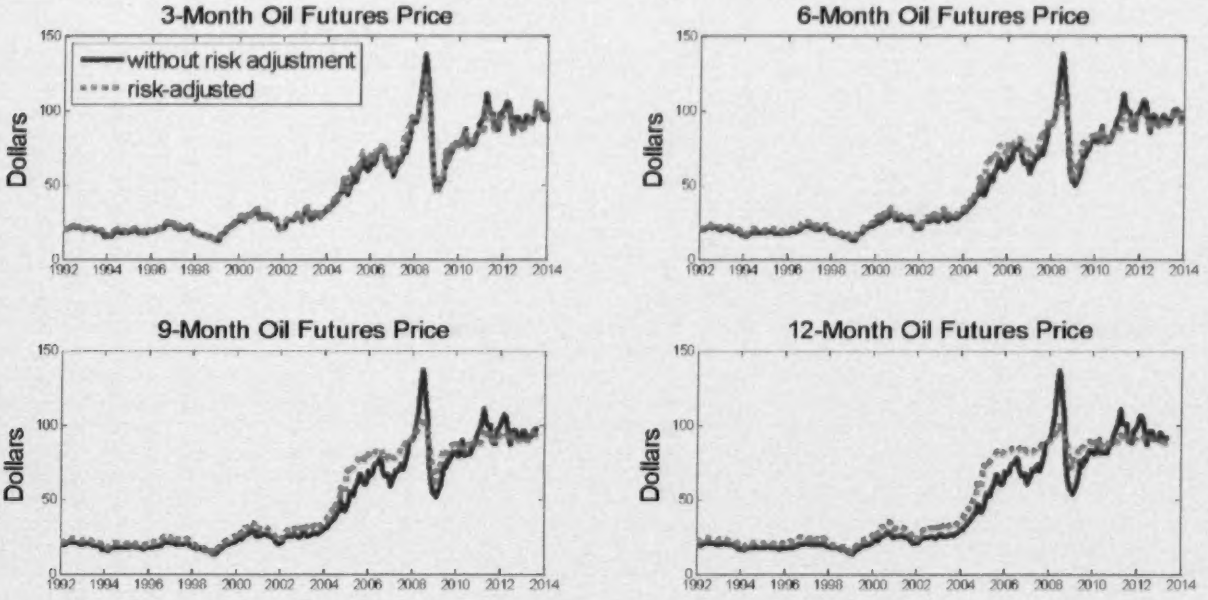
NOTES: The plot shows 17 alternative estimates of the time-varying risk premium based on predictors that have been proposed in the related literature. The plot illustrates that there is substantial disagreement on the magnitude and sign of the time-varying risk premium. Qualitatively similar results are obtained for other horizons.

Figure 2: Oil Price Expectations Based on the Payoff Regression with All Predictors



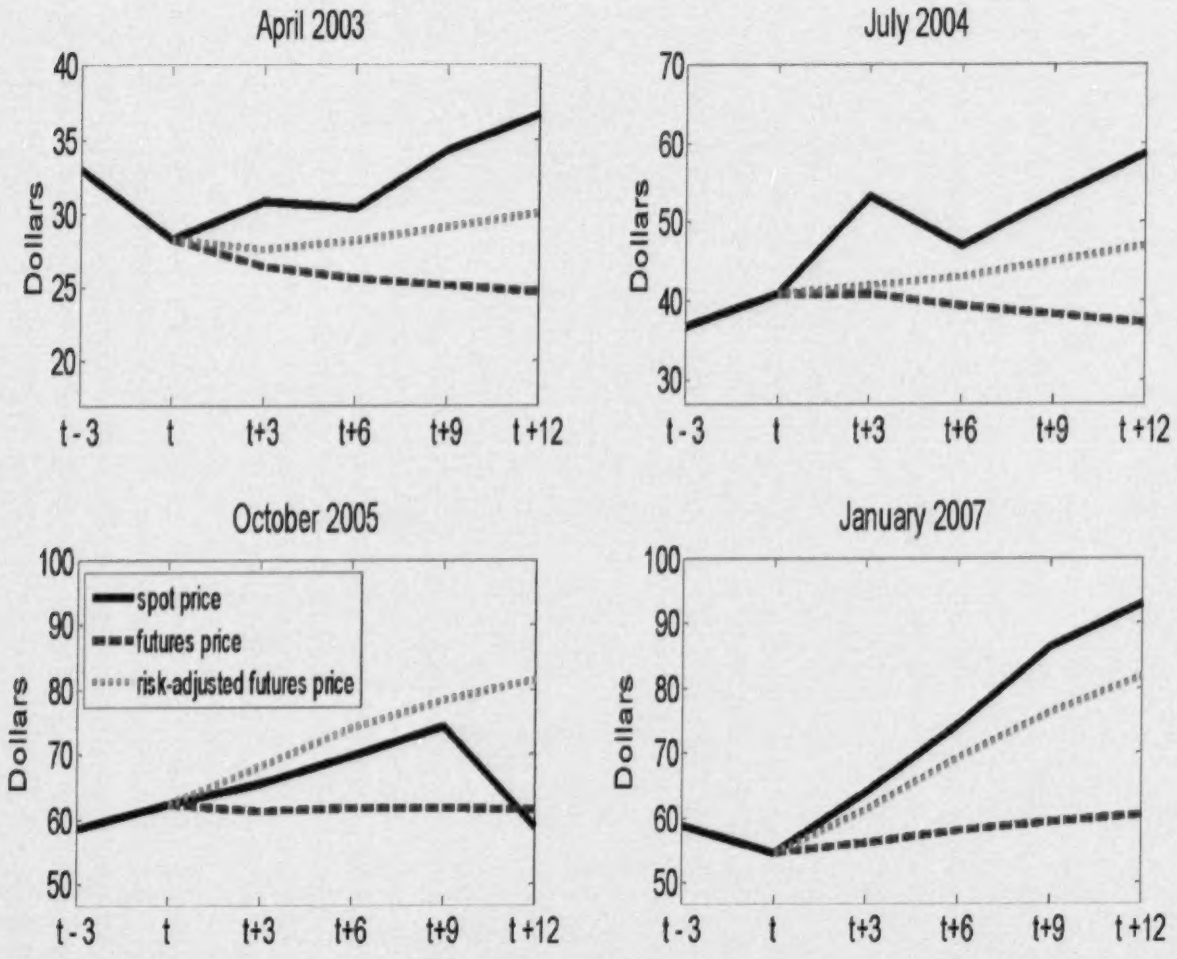
NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an unrestricted payoff regression including all 30 predictors listed in Table 5.

Figure 3: Oil Price Expectations based on the HW Term-Structure Model



NOTES: The risk-adjusted futures price is a measure of the futures market's expectation of the price of oil. The risk adjustment is derived from an updated estimate of the term structure model in Hamilton and Wu (2014).

Figure 4: Selected trajectories of F_t^h, the Realized Spot Price S_{t+h}, and $F_t^h - \widehat{RP}_t^h$, from HW Model



NOTES: The estimates of the time-varying risk premium are based on the term structure model of Hamilton and Wu (2014) updated to June 2014. The risk-adjusted futures price is a measure of the market's oil price expectation at horizons of 3, 6, 9, and 12 months. The plot illustrates that adjusting the oil futures price for the HW estimate of the time-varying risk premium diminishes the systematic prediction errors that arise when using oil futures prices as proxies for oil price expectations.

Appendix

This appendix reports additional results for monthly and quarterly models to demonstrate the robustness of our main findings. It also discusses the data sources and the construction of the variables.

Table A.1. Risk-Adjusted Out-of-Sample Forecasts of the Nominal WTI Spot Price Based on Monthly Recursive Estimates of the Risk Premium MSPE Ratio for Evaluation Period: 1992.1-2014.6

Horizon h	F_t^h	B1	B2	BC	S	DNV1	DNV2	DNV3
3	**0.987**	1.039	1.070	1.252	1.256	1.032	1.114	1.034
6	**0.982**	1.117	1.137	1.414	1.335	1.153	1.240	1.161
9	**0.949**	1.132	1.150	1.646	1.209	1.249	1.348	1.270
12	**0.882***	1.125	1.128	2.121	1.422	1.202	1.476	1.214

Horizon h	GHR1	GHR2	HY1	HY2	PP1	PP2	PP3	BS
3	1.092	1.116	1.117	1.133	1.093	1.115	**0.985***	-
6	1.217	1.244	1.244	1.226	1.216	1.264	1.093	-
9	1.286	1.328	1.389	1.398	1.336	1.449	1.145	-
12	1.153	1.450	1.658	1.846	1.444	1.753	1.138	-

NOTES: See Table 8. The evaluation period for B1, B2, BC and S at h=3 ends in 2013.12 due to data constraints. The initial estimation period for BS is too short for recursive estimation; therefore, this model is excluded. All variable transformations are performed only with the data available at each point in time. To obtain the normalized crude oil inventories a one-sided HP-filter with lambda=160,000 is used.

Table A.2. Quarterly Predictor Variables for Payoff on Oil Futures Contracts

Article	Model	Predictors
Etula (2013)	E1	Effective risk aversion (broker-dealer variable) S&P500 excess return (proxy for market risk)
	E2	Effective risk aversion (broker-dealer variable)
	E3	Effective risk aversion (broker-dealer variable) VIX implied volatility for S&P500 3-month T-bill rate Yield spread (Aaa – 3-month T-Bill) Dividend yield on S&P 500 common stock portfolio U.S. CPI inflation
	E4	E3 + basis + own-market hedging pressure
Acharya, Lochstoer and Ramadorai (2013)	ALR1	Expected default frequency (EDF) in oil & gas industry Basis Default spread (Baa – Aaa) Median SPF forecast of quarterly GDP growth 3-month T-bill rate
	ALR2	Zmijewski-score (Zm) of default risk in oil & gas industry Basis Default spread (Baa – Aaa) Median SPF forecast of quarterly GDP growth 3-month T-bill rate
	ALR3	ALR1 + Realized quarterly variance of oil futures returns (RV) + Interaction between EDF and RV
	ALR4	ALR2 + Realized quarterly variance of oil futures returns (RV) + Interaction between Zm and RV
	ALR5	ALR1 + Effective risk aversion (broker-dealer variable) + Interaction between EDF and effective risk aversion
	ALR6	ALR2 + Effective risk aversion (broker-dealer variable) + Interaction between Zm and effective risk aversion

NOTES: The sample period is 1986.1-2014.2 except for the EDF series and the Zmijewski-score which are only available until 2009.4 and 2010.4, respectively.

Table A.3. Predictive Accuracy of Risk-Adjusted Futures Prices Based on Quarterly Full-Sample Estimates

MSPE Ratio for Evaluation Period: 1992.I-2014.II

Horizon h	F_t^h	HW	E1	E2	E3	E4
1	0.984**	0.822*	1.061	1.048	0.953	0.908
2	0.983	0.685*	1.107	1.108	0.915**	0.911**
3	0.954**	0.587*	1.119	1.118	0.862*	0.828*
4	0.905*	0.526*	1.110	1.114	0.761*	0.748*

Horizon h	ALR1	ALR2	ALR3	ALR4	ALR5	ALR6	All predictors
1	0.834**	0.827**	0.782**	0.801**	0.751**	0.758**	0.711
2	0.855**	0.880*	0.811**	0.862*	0.761**	0.794*	0.754**
3	0.767**	0.834*	0.741**	0.833*	0.629**	0.709*	0.607*
4	0.707**	0.671*	0.668*	0.664*	0.668*	0.623*	0.513*

NOTES: See Table 4. The evaluation period for ALR1, ALR3, and ALR5 ends h periods after 2009.IV and for ALR2, ALR4, and ALR6 ends h periods after 2010.IV due to data constraints. Adjusting the evaluation period underlying the MSPE ratios for F_t^h accordingly produces results qualitatively similar to the results shown in Table A.3.

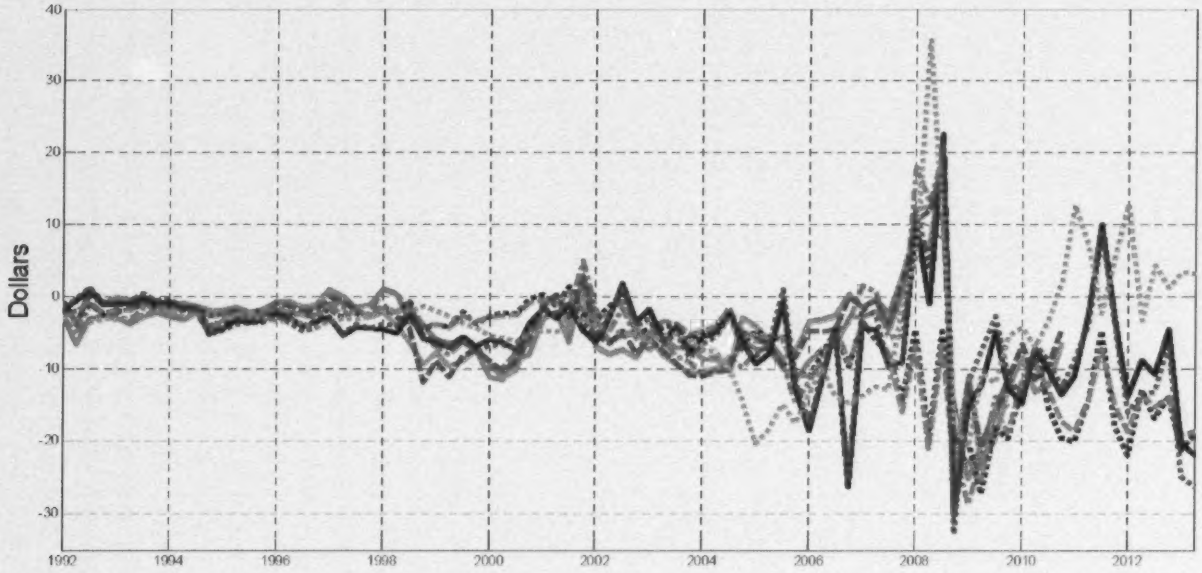
Table A.4. Risk-Adjusted Out-of-Sample Forecasts of the Nominal WTI Spot Price Based on Quarterly Recursive Estimates
MSPE Ratio for Evaluation Period: 1992.I-2014.II

Horizon h	F_t^h	HW	E1	E2	E3	E4
1	0.984**	1.168	1.181	1.138	1.242	1.240
2	0.983	1.187	1.268	1.243	1.381	1.427
3	0.954**	1.310	1.311	1.295	1.555	1.587
4	0.905*	1.470	1.410	1.386	-	-

Horizon h	ALR1	ALR2	ALR3	ALR4	ALR5	ALR6
1	1.093	1.056	1.097	1.064	1.177	1.137
2	1.191	1.137	1.238	1.169	1.514	1.203
3	1.243	1.159	1.344	1.224	2.107	1.203
4	-	-	-	-	-	-

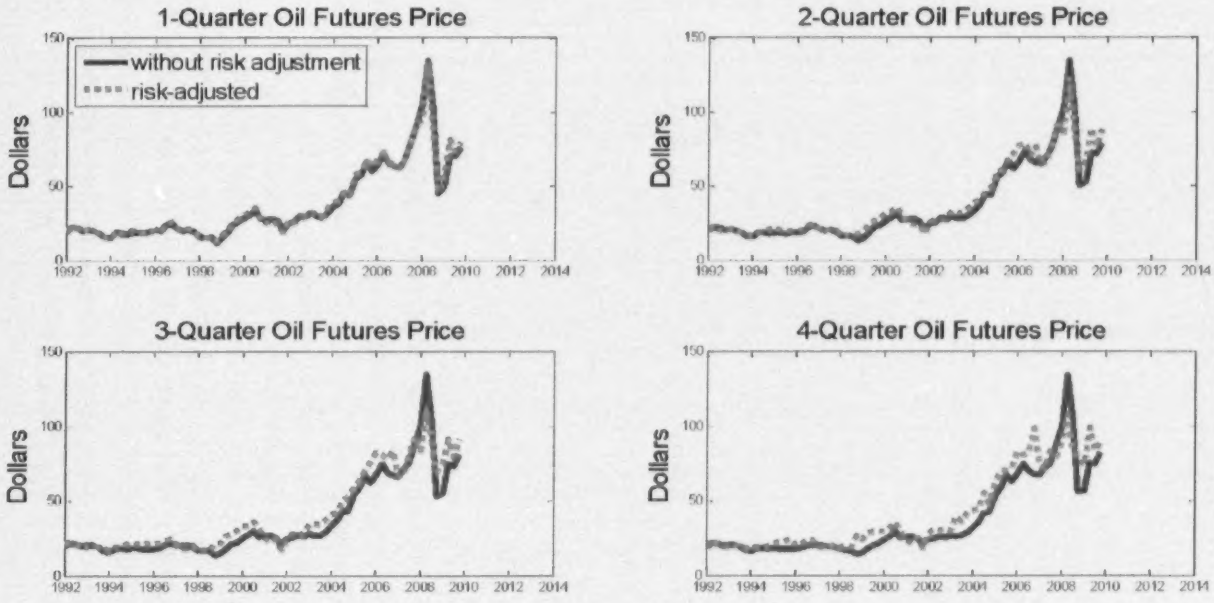
NOTES: See Table 4. All MSPE ratios have been normalized relative to the monthly no-change forecast. The evaluation period for ALR1, ALR3, and ALR5 ends h periods after 2009.IV and for ALR2, ALR4, and ALR6 ends h periods after in 2010.IV due to data constraints. Adjusting the evaluation period underlying the MSPE ratio s for F_t^h accordingly produces results qualitatively similar to the results shown in Table A.4. We only report results for the payoff regressions if $K/T \leq 0.5$, where K is the number of regressors and T is the initial recursive sample size.

Figure A.1: Alternative Quarterly Estimates of the Time-Varying Risk Premium in the Oil Futures Market at the 4-Quarter Horizon



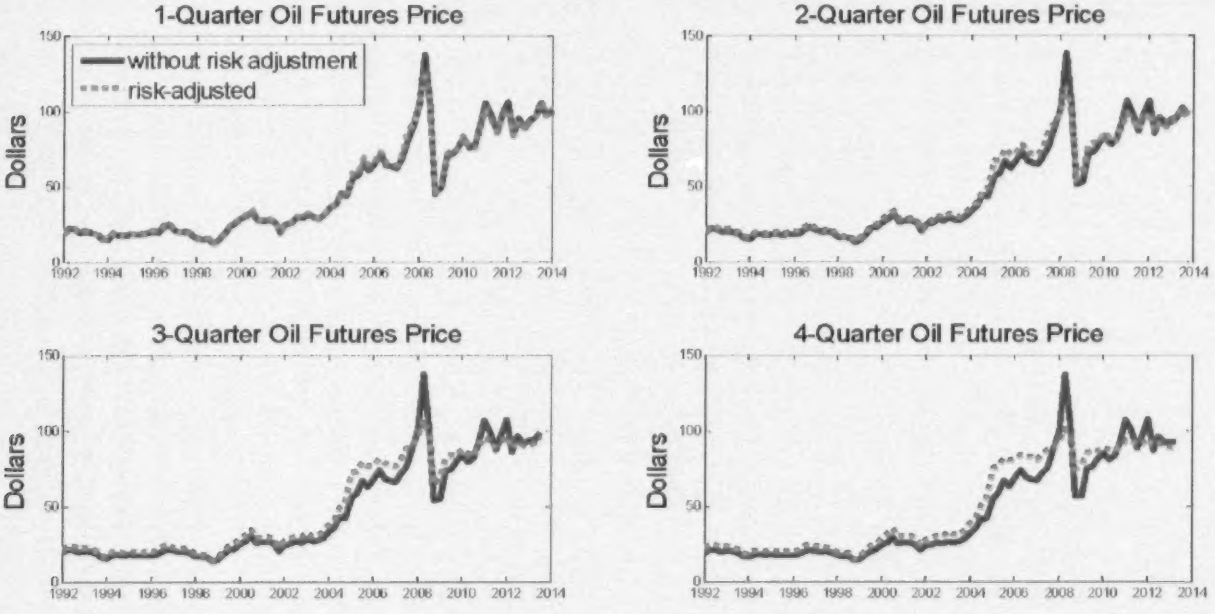
NOTES: The plot shows the 11 alternative estimates of the quarterly time-varying risk premium proposed in the literature, as summarized in Table A.2. The plot illustrates that there is substantial disagreement on the magnitude and sign of the time-varying risk premium. Qualitatively similar results are obtained for other horizons.

Figure A.2: Oil Price Expectations based on the Quarterly All-Predictor Regression



NOTES: The risk-premium adjustment is derived from an unrestricted payoff regression including all 19 predictors listed in Table A.2. The evaluation period is restricted due to the availability of some of the quarterly predictors.

Figure A.3: Oil Price Expectations based on the HW Term Structure Model



Data Sources and Construction of the Variables

Monthly averages of the daily WTI spot price were obtained from the U.S. Energy Information Administration (EIA) starting in 1986.1. The corresponding daily oil futures prices for maturities between 1 and 12 months are from Bloomberg going back as far as April 1983 depending on the maturity.

A. Macroeconomic and financial data

Monthly data for the 1-month T-bill rate, the 3-month T-bill rate, constant maturity rates for Treasury government bonds with maturities 1, 2, 5, 10 and 20 years, and Moody's Aaa and Baa corporate bond yields were obtained from the St. Louis Fed FRED database. From the same database we also retrieved data on the degree of capacity utilization in U.S. manufacturing measured in percent, U.S. consumer prices for all urban consumers and U.S. industrial production. Following Bessembinder (1992), annualized U.S. CPI inflation is decomposed into an expected and an unexpected component. Expected inflation is defined as the difference between the 1-month T-bill rate and the expected real interest rate which is determined as the simple average of ex-post realized real Treasury-bill returns over the preceding 12 months. Unexpected inflation is the difference between actual and expected inflation. The unexpected change in U.S. industrial production is computed based on the residuals from an ARIMA(3,1,0) model fitted to the raw industrial production data following Bessembinder (1992). The composite leading indicator for the OECD economies and six non-OECD economies (Brazil, China, India, Indonesia, Russia and South Africa) was obtained from the OECD *Main Economic Indicator* database. We follow Pagano and Pisani (2009) in selecting the ratio-to-trend series which expresses the cyclical behavior of the indicator as deviations from its long-run trend.

Returns on the CRSP value-weighted equity index with and without dividends are available via the Wharton Research Data Services. The dividend yield is constructed based on those two return series as described in Fama and French (1988). The market portfolio excess return used in Sadorsky (2002) is the equally weighted return on the CRSP value-weighted equity index (excluding dividends) and the Dow Jones-AIG Commodity Index in excess of the 3-month T-bill rate. The Dow Jones-AIG Commodity Index is composed of futures contracts on 19 physical commodities and is obtained from Bloomberg. Data for the S&P 500 stock price index and the dividend yield on the S&P 500 are from Haver Analytics. The VIX volatility index is obtained from Bloomberg.

The median 1-quarter-ahead forecast of U.S. real GDP growth from the *Survey of Professional Forecasters* is available from the Philadelphia Fed website. Etula's (2013) measure of effective risk aversion is based on balance sheet data for total financial assets and liabilities of U.S. security brokers and dealers as well as U.S. households and non-profit organizations and is obtained from the Federal Reserve's Flow of Funds database. The effective risk aversion index is constructed as: 1 + broker-dealer equity/household equity × (1 – broker-dealer leverage/market leverage), where equity is defined as the difference between financial assets and liabilities and leverage is computed as total assets over equity and the "market" refers to both U.S. security broker-dealers and households. A secular downward trend is removed by subtracting the 4-quarter moving average. The quarterly time series for the expected default frequency in the U.S. oil and gas industry and for the Zmijewski-score of default risk of U.S. oil and gas producers developed by Acharya, Lochstoer and Ramadorai (2013) are available on the authors' website (http://pages.stern.nyu.edu/~sternfin/vacharya/).

B. Commodity-specific data

Monthly data for U.S. crude oil commercial inventories excluding strategic petroleum reserves are reported in the *Monthly Energy Review* issued by the EIA. The normalized oil inventories are obtained by deseasonalizing and detrending the log of inventories using a Hodrick-Prescott filter with a smoothing parameter of 160,000 following Gorton, Hayashi and Rouwenhorst (2013).

Data on long and short positions of commercial traders for the crude oil, heating oil, gold, silver and platinum futures markets and open interest for the oil futures market are obtained from the reports on the *Commitment of Traders in Commodity Futures* published by the CFTC. Hedging pressure is defined as (number of short commercial positions – number of long commercial positions)/total number of commercial positions) for each market. Price pressure is the month-on-month change in hedging pressure. Following Hong and Yogo (2012) the oil market dollar open interest is computed as the total number of futures contracts outstanding times the spot price of crude oil; given that the monthly growth rate of this variable is noisy, they smooth it by taking the 12-month geometric average. Daily data on open interest and volume of trades of oil futures contracts for maturities 3, 6, 9 and 12 months are obtained from Bloomberg. The monthly liquidity measure is constructed as the median of the daily ratios of trading volume over open interest for each horizon. We follow Hong and Yogo (2012) in computing the basis as $(F_t^h / S_t)^{(1/h)} - 1$ for their return regression, whereas for the Fama and French regressions we define it as the percent spread between F_t^h and S_t without adjusting for the maturity, as in their original work.